Latent Variables, Feature Importance and Regression
Session 4

Dr. Ali Arsanjani
@AliArsanjani
Adjunct At San Jose State University
Senior Lecturer UCSD

ML in the News: social impact of ML

The Social Dilemma

PG-13 2020 · Documentary · 1h 33m

Watch now
Subscription

Watched it?

Watchlist

All watch options

8.1/10	87%	78%
IMDb	Rotten Tomatoes	Metacritic

97% liked this movie

Google users

Micro-segmentation for hyper-personalization

- Process :

1. EDAV
   1. Visualization
2. Cluster →
   1. Micro-clustering
3. (add a column with a label) → supervised learning → Classification →
4. select subset, categorical → numerical →
5. Regression → Prediction →
6. add context → decision
7. → action

Regression in context

Un-supervised | supervised

cluster

Logistic regression
classify

Linear regression
Predict

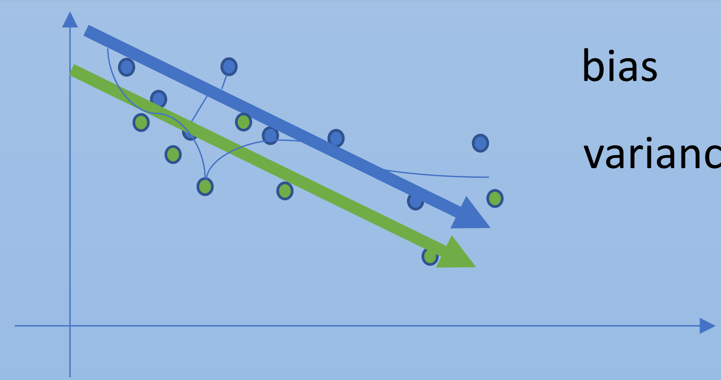
Build context

Recommend

Forecasting

$Y = wx + b$

$Y = a[i]\,x[i] + a[0]$



bias

latent

variance

(c) Ali Arsanjani, 2019

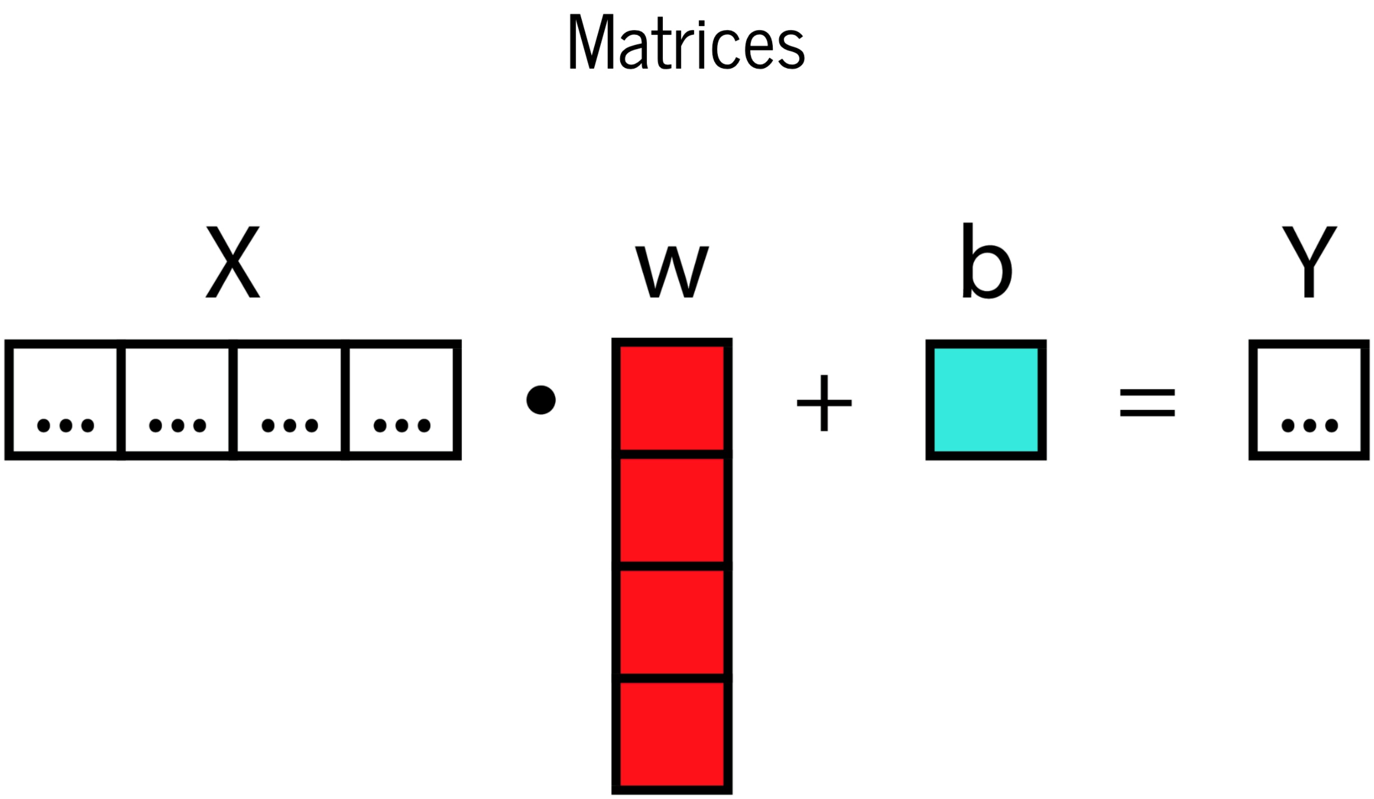
Linear Regression

Function

$$Y = f(X, w_1, b_1)$$

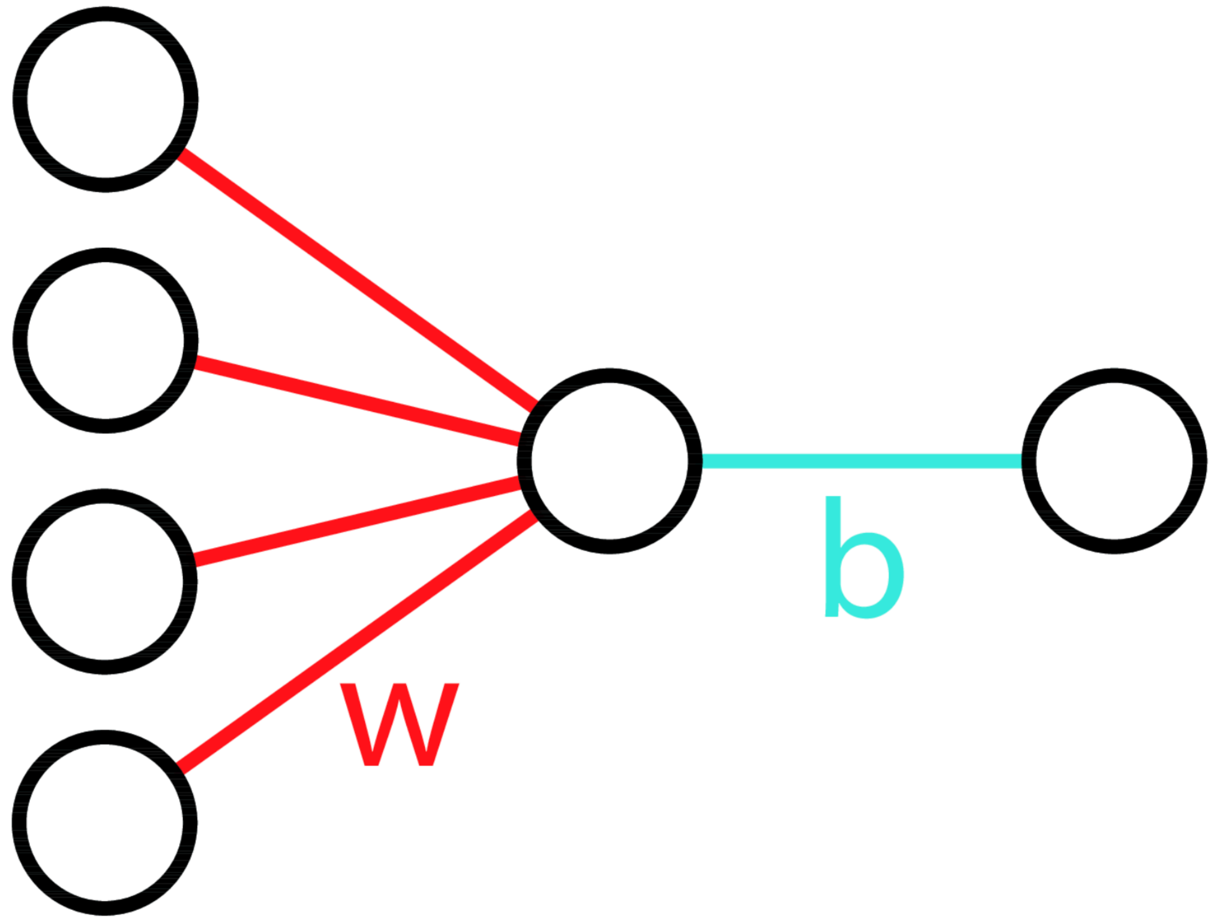
$$f(x, w, b) = x \cdot w + b$$

In Matrix Format



Source: ericmjl/bayesian-deep-learning-demystified

Neural Diagram

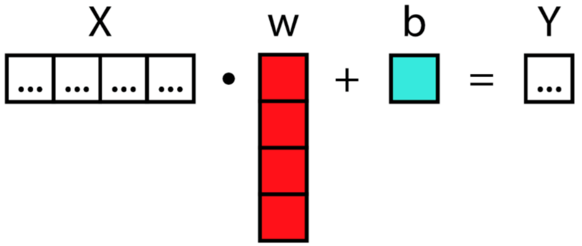


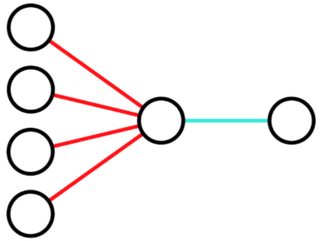
Source: ericmjl/bayesian-deep-l

LinReg 3 Ways

$$Y = f(X, w_1, b_1)$$

$$f(x, w, b) = x \cdot w + b$$





Source: ericmjl/bayesian-deep-learning-demystified

Logistic Regression is a binary classification solution , linear regression is a prediction (number) solution

- Logistic is a special form of linear regression
- $1/(1-e^x)$

Logistic Regression

Function

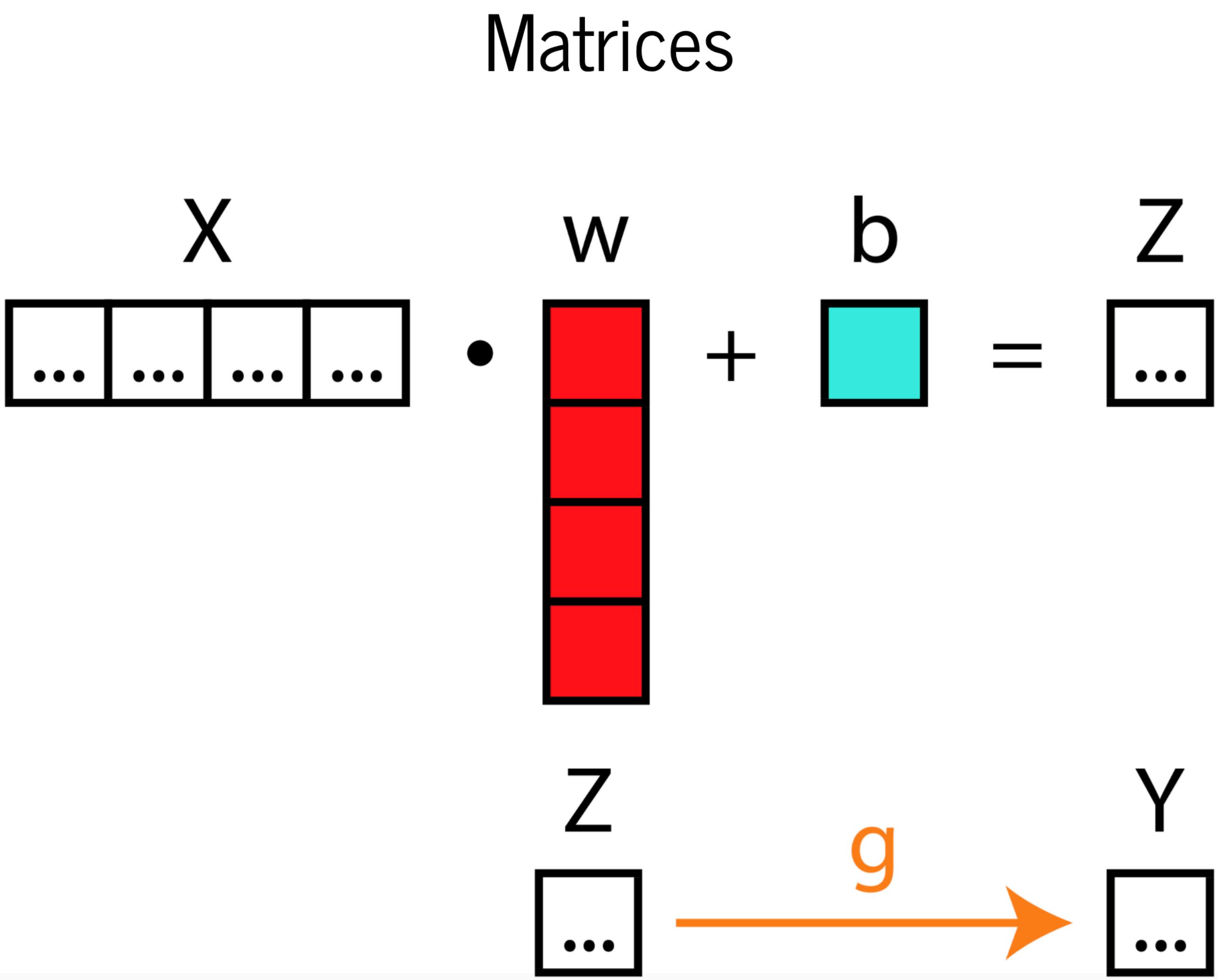
$$Z = f(X, w_1, b_1)$$

$$Y = g(Z)$$

$$g(x) = \frac{1}{1+e^{-x}}$$

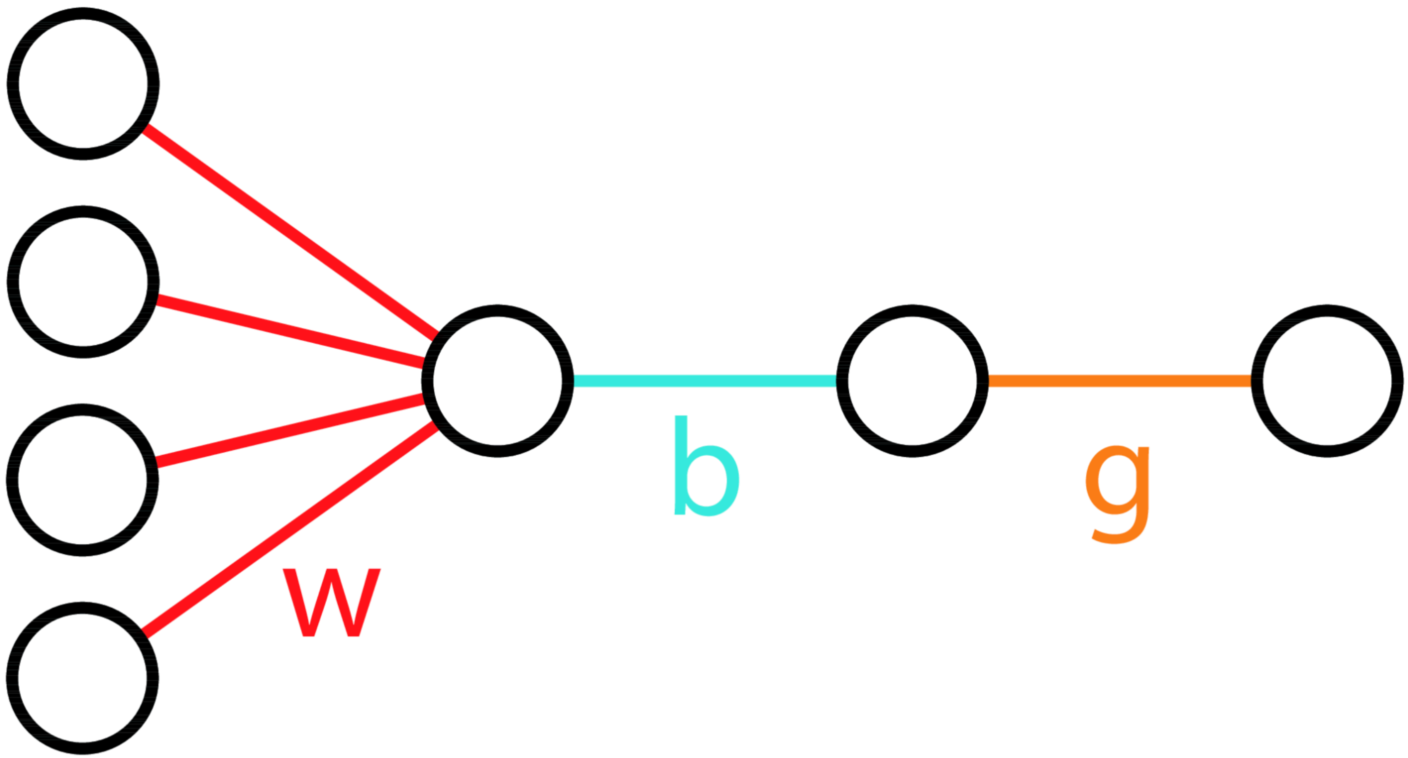
Source: ericmjl/bayesian-deep-learning-demystified

Matrix View of Logistic Regression



Source: ericmjl/bayesi

Neural Diagram

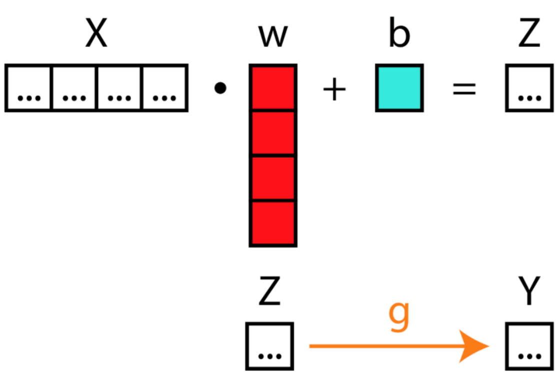


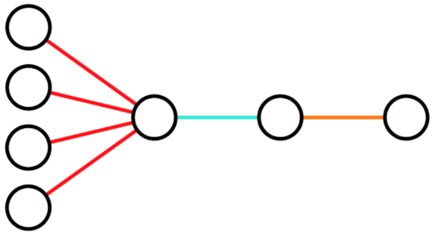
Source: ericmjl/bayesian deep learning demystified

LogReg 3 Ways

$$Z = f(X, {\color{red}w_1}, {\color{blue}b_1})$$
$$Y = g(Z)$$

$$g(x) = \frac{1}{1 + e^{-x}}$$





Source: ericmjl/bayesian-deep-learning-demystified

Deep Neural Networks

Function

$$Z = \tanh(f(X, w_1, b_1))$$

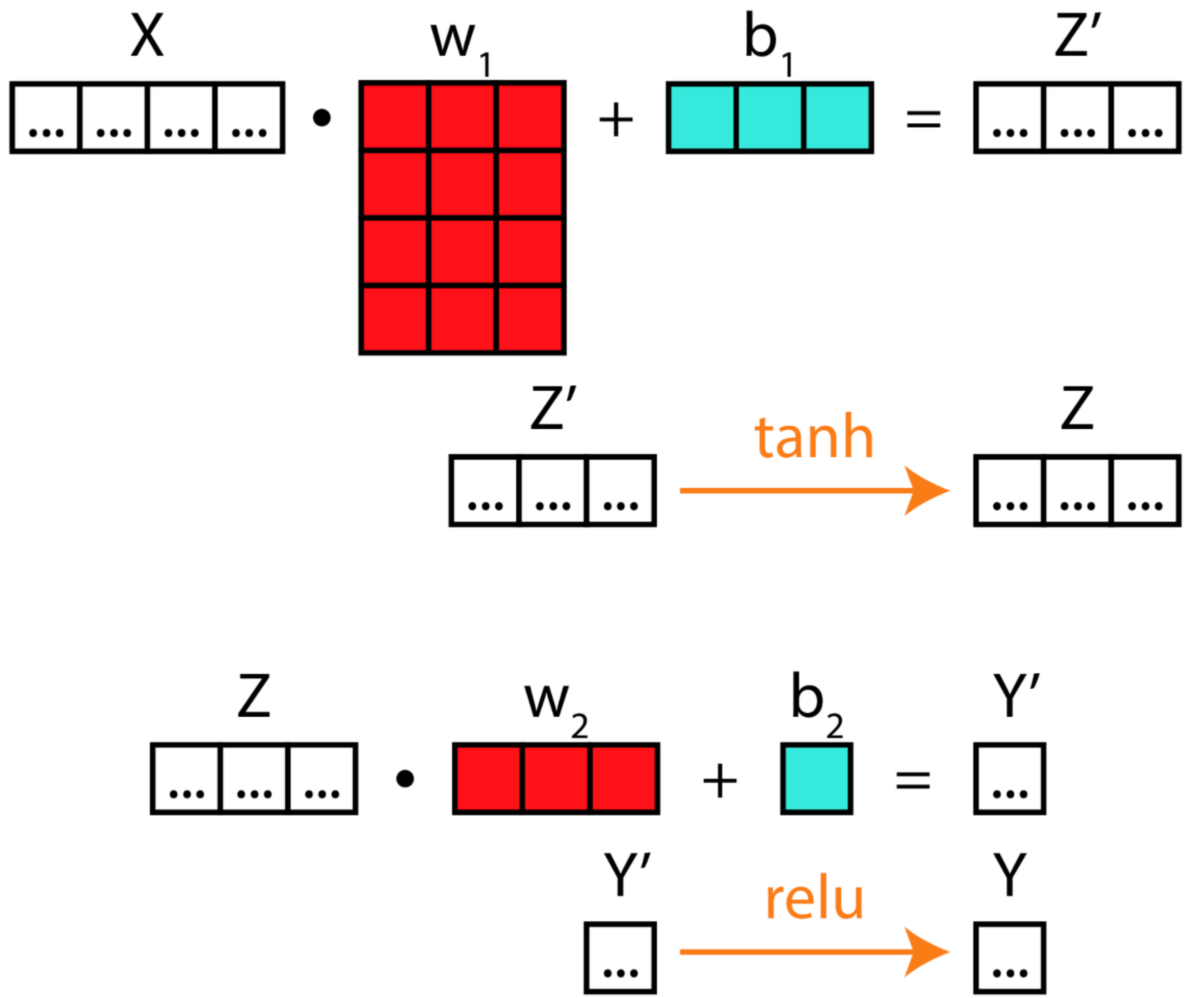
$$Y = \mathrm{ReLu}(f(Z, w_2, b_2))$$

$$\mathrm{ReLu}(x) = \max(x, 0)$$

$$\tanh x = \frac{e^{2x} - 1}{e^{2x} + 1}$$

Source: ericmjl/bayesian-deep-learning-demystified

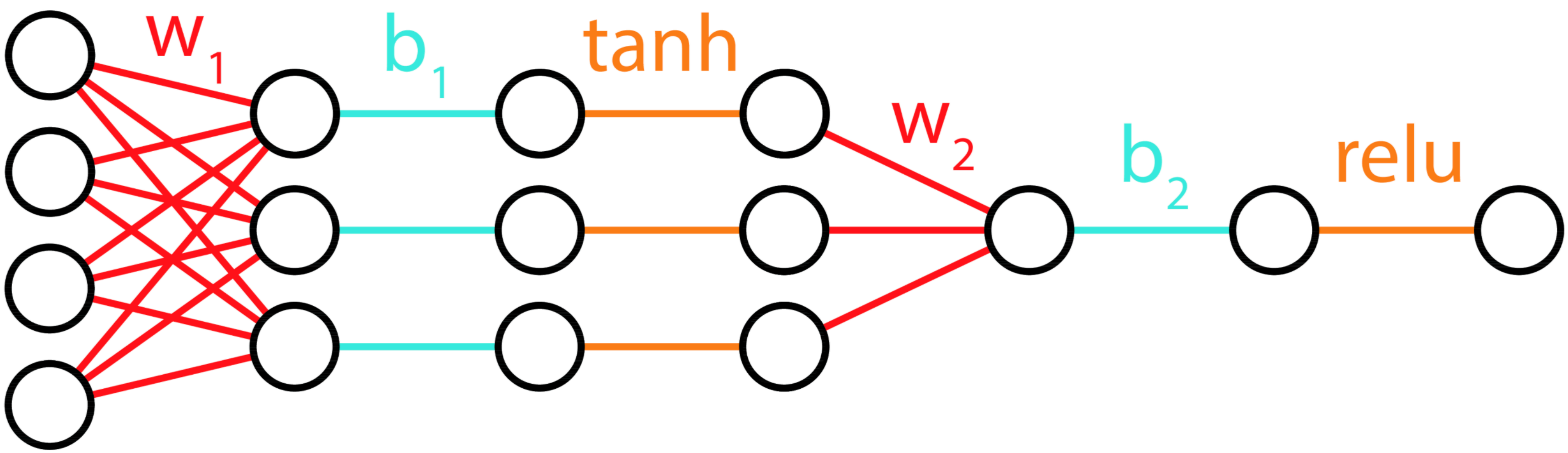
Matrices



Source: ericmjl/bayesian-deep-learning-demystified

Let’s take a Neural View

Neural Diagram



Source: ericmjl/bayesian-deep-learning-demystified

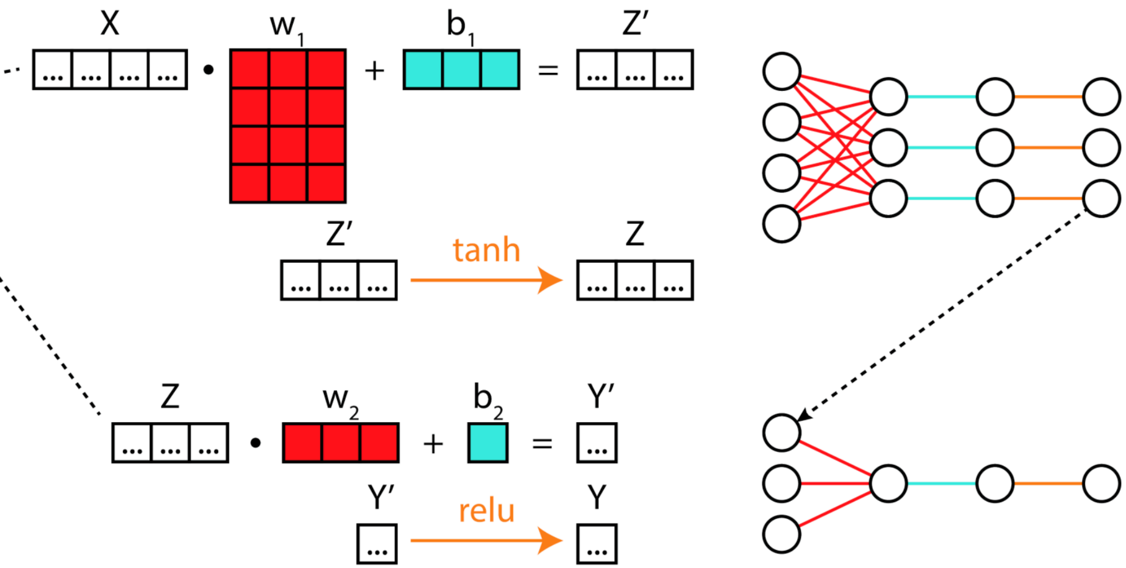
DeepNets 3 Ways

$$Z = \tanh(f(X, {\color{red}w_1}, {\color{blue}b_1}))$$

$$Y = \mathrm{ReLu}(f(Z, {\color{red}w_2}, {\color{blue}b_2}))$$

$$\mathrm{ReLu}(x) = \max(x, 0)$$

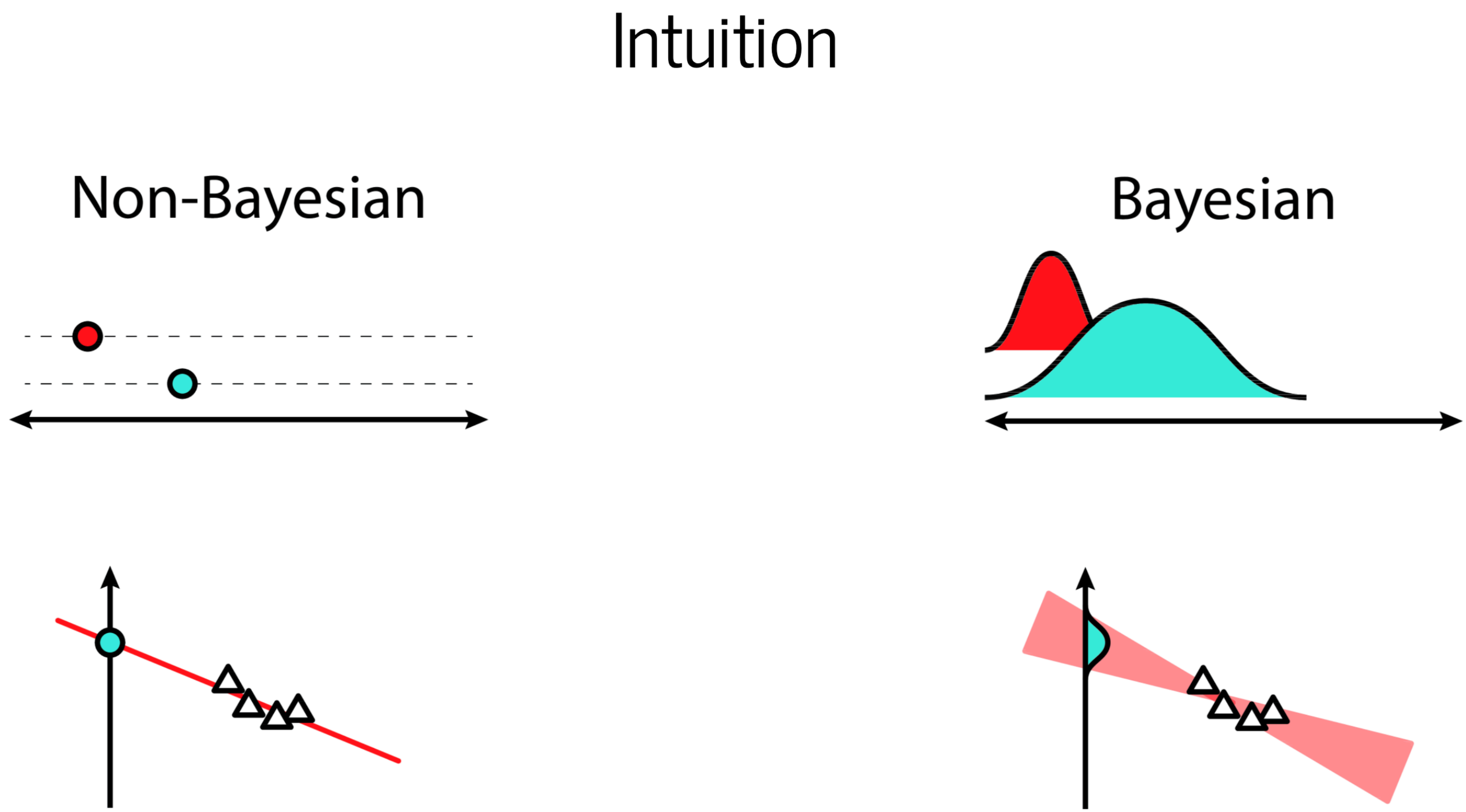
$$\tanh x = \frac{e^{2x} - 1}{e^{2x} + 1}$$



Source: ericmjl/bayesian-deep-learning-demystified

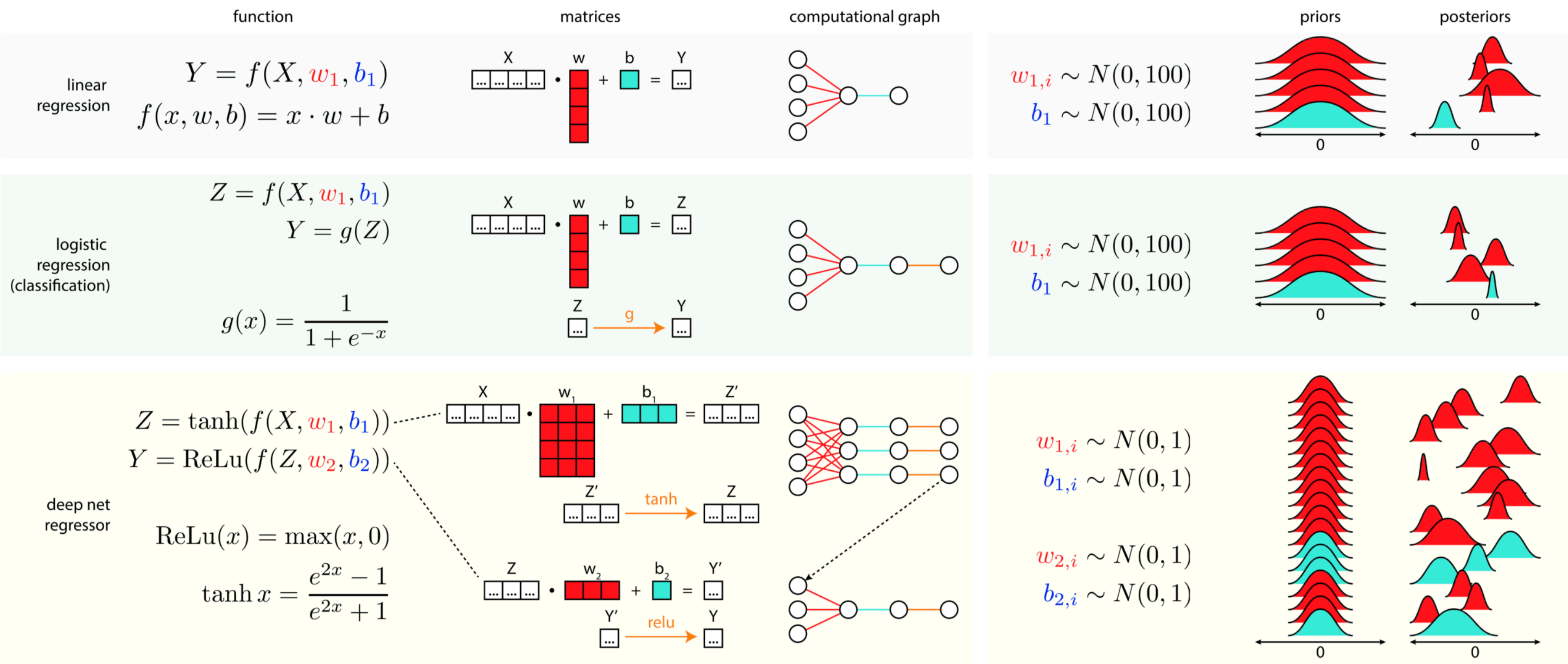
Bayesian Neural Networks

Key Idea: Learn probability density over parameter space.



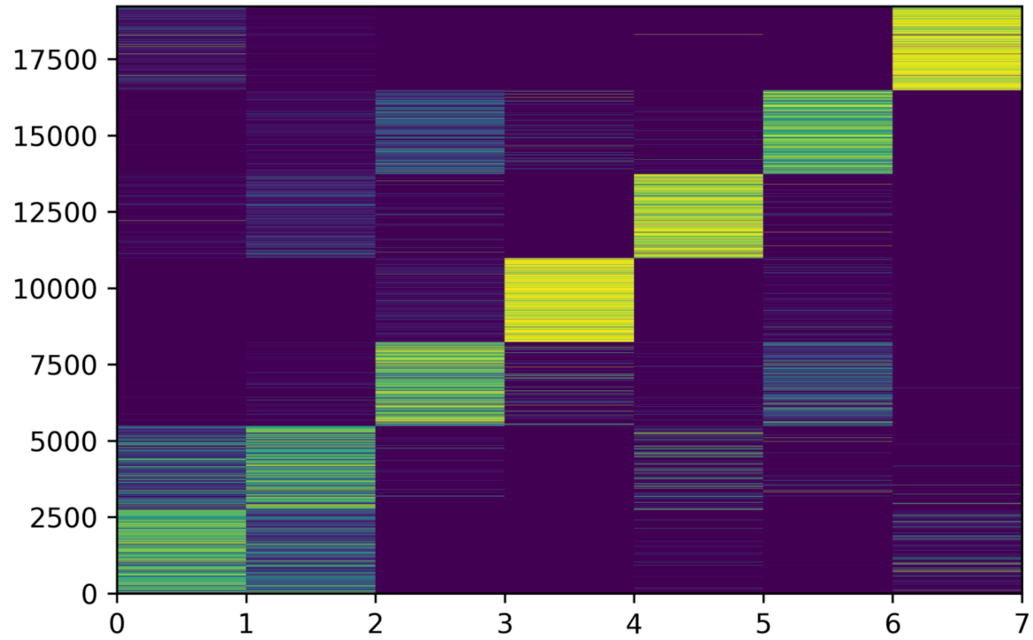
Source: ericmjl/bayesian-deep-learning-demystified

Cheat Sheet



Source: ericmjl/bayesian-deep-learning-demystified

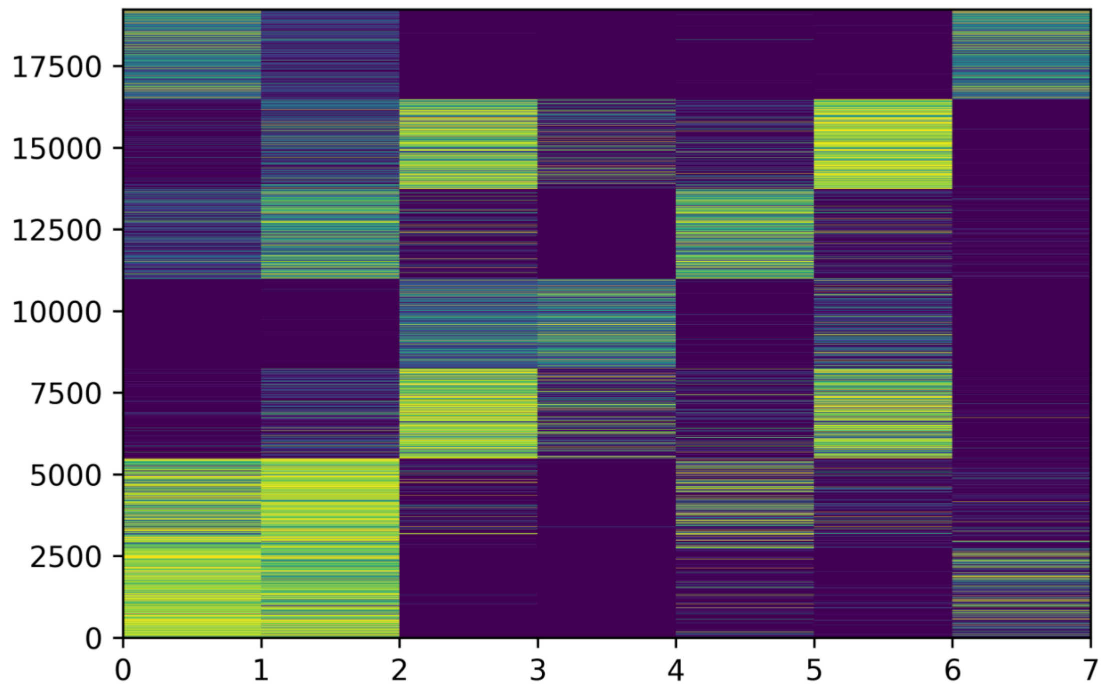
Predict Forest Cover Type
Class Probabilities



"probabilistic estimate"

Source: ericmjl/bayesian-deep-learning-demystified

Predict Forest Cover Type
Class Uncertainties

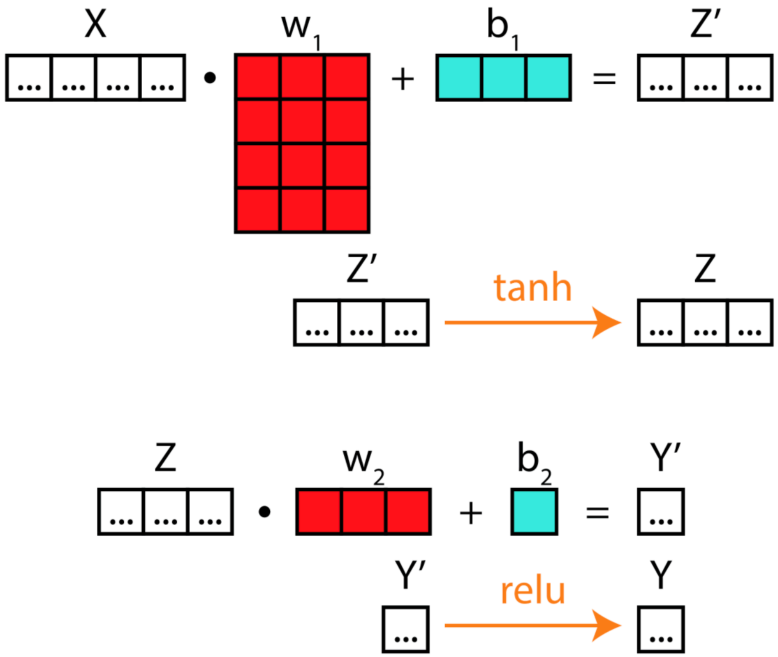


"with uncertainties!"

Source: ericmjl/bayesian-deep-learning-demystified

Take-Home Point 1

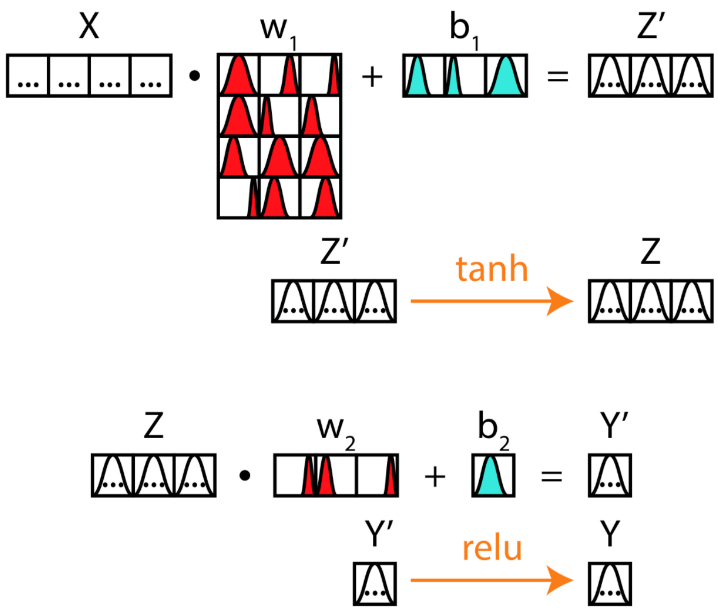
Deep Learning is nothing more than **compositions of functions on matrices.**



Source: ericmjl/bayesian-deep-learning-demystified

Take-Home Point 2

Bayesian deep learning is grounded on **learning a probability distribution for each parameter.**



Source: ericmjl/bayesian-deep-learning-demystified

Location as a latent variable / latent manifold in the prediction of house prices

Num tenants	age
	50

rooms	Sq footage	bathrooms	Address (zipcode)		Computed location index value
3	1500	1.5	12345		

DataSet1

Rmse = r1 >> 0

Hidden variables

Proximity to schools	Prox to shopping	Prox to highway	Prox to rest	walkability	zipcode
					12345

DataSet2 Location

Gini/ f imp

Latent Space

Loc index for buyers= .2s+ .4sh + .2h + .2r

Loc index for renters= .001s+ .4sh + .2h + .05r .7w

Use-case : Stock Clustering

- Objective function → cluster types (understand the significance of the clusters; the meaning behind each cluster each segment)
  - Highest gain clusters
  - Consider other functions (objective) , label the new clusters
    - One dataset 2 additional columns one for each cluster type
      - Highest annual revenue, Lowest variance, f(x) + g(x) = h(x)
      - cluster on the one composite function
- Standards for labeling
  - Categorical , integer (1,2,3,) , top, middle, lower ➔ numbers (categorical mapping) CountVectorizer in scikit learn

In class Group task: 6:36 – 6:55 pm

- Huddle on Latent Variable takeaways from Chopra Paper
- Re-address your project
- Be prepared to present it
- And talk about your fractal clustering

Chopra Paper Takeaways

- Ensemble model
- Latent manifold
- Data Amalgmation
- Read abstract, intro, conclusion, then scan titles and look for eg latent variable section.


Deep
Context
AI